T.A. Malysheva, O.V. Anchugova
Ural Federal University named after the first President of Russia B.N. Yeltsin
Yekaterinburg, Russia

USING EXTRACTED DATA FROM SOCIAL NETWORKS

Abstract: Every day a significant part of the world population spends several hours on social networks. People fill information in their accounts about themselves, regularly updating it, thereby, they create a huge up-to-date database, which can be extracted and analyzed by someone. Such extraction often happened without informing people whose data were extracted. This can be a threat to the personal data of social networks' users. However, users' data can also be used for good purposes, such as to improve existing business processes and social ones, which ultimately can benefit society and can help automate many processes. These aspects will be discussed in this article. The article considers some options for using data from social networks, such as identifying stress based on social interactions in social networks, creating a reliable online survey using personal information from social networks, automated data collection for calculating credit ratings based on financial transactions and social networks, and also features of the Russian legislation in the field of personal data.

Keywords: data, social networks, social media, social interaction, personal information.

Т.А. Малышева, О.В. Анчугова
Уральский федеральный университет имени первого Президента России Б.Н. Ельцина
Екатеринбург, Россия

ИСПОЛЬЗОВАНИЕ ИЗВЛЕЧЕННЫХ ДАННЫХ ИЗ СОЦИАЛЬНЫХ СЕТЕЙ

Аннотация: Каждый день значительная часть населения земли тратит несколько часов на социальные сети. Люди наполняют свои аккаунты информацией о себе, регулярно обновляя ее, тем самым создают огромную актуальную базу данных, которую кто-то сможет извлечь и проанализировать. Такое извлечение зачастую может происходить без ведома людей, чьи данные извлекаются, и в корыстных целях. Это может представлять угрозу персональным данным пользователям социальных сетей. Тем не менее, данные пользователей социальных сетей также могут использоваться и для благих целей, для усовершенствования существующих как бизнес процессов, так и социальных, что в итоге может принести благо в общество и автоматизировать множество процессов, которые не были автоматизированы ранее. В данной статье будут рассмотрены некоторые варианты использования данных из социальных сетей, такие как: выявление стресса на основе социальных взаимодействий в социальных сетях, формирование достоверного онлайн-опроса с использованием персональной информации из социальных сетей, автоматизированный сбор данных для расчета кредитного рейтинга на основе финансовых операций и социальных сетей, а также особенности российского законодательства в области персональных данных.

Ключевые слова: данные, социальные сети, социальное взаимодействие, персональные данные.

Introduction

According to Hootsuite statistics (January 2018), more than 42% of the world population are active users of social media and 39% are active mobile social users. If we consider the developed countries separately, these indicators will be much higher.

This topic is relevant, as social networks have become part of everyday life of a modern person. Today social nets contain a lot of relevant information that can be used for different purposes.

The purpose of this article is to consider possible options for using extracted data from social networks, as well as the legal basis for processing data from social networks in Russia.

The purpose of the article defines the following research questions:

- to give a brief description of the extracted data;
- to consider options for using a dataset from social networks;
- to study Russian legislation in the field of using personal data from social networks and describe the problem of data security.

The method of research in this article is the analysis of scientific articles and legislative acts.

Data from social networks

At the moment the most popular social networks in Russia are: Vkontakte, Instagram, Facebook and Odnoklassniki.

Undoubtedly, the personal information contained in social networks may be of interest to third parties. Such information can be divided into three types: Profile data, Social graph and Traffic data.

Profile data. This information includes personal data of users which they upload to their profiles by themselves. Users voluntarily download this data but most of them expect that only friends will see this information. Profile data can include name, surname, date of birth, city of residence, education information, place of work, photos, personal preferences, interests, etc. Also, we can include information about the profile change history that shows the chronology of the user's activity.

Social graph. The peaks of this graph show users. The borders show their friendly relations. Extracting social graphs carries a greater threat, since it is easier to extract a social graph than a complete set of a user profiles. This graph has a wide range of its uses: identifying user communities with similar interests, identifying well connected persons, receiving personal information from friends of a user, etc.

Traffic data. Most social networks collect and store data about the user's location, IP address, used browser, as well as the duration and frequency of sessions. This type of information represents the least danger, since third parties are not likely extract such information.

Options of using a dataset from social networks

Many authors consider the possibility of using personal data from social networks, creating systems that improve or make person's life easier or optimize business processes. Further we compare some of them.

For example, in the article «Detecting Stress Based on Social Interactions in Social Networks» [2], the authors say that the stress state of users is closely related to the state of his social friends. Psychological stress threatens the health of people. It is easy to detect stress in a timely manner, for an early care. With the popularity of social networks, people get used to sharing their daily activities and interacting with friends on social networking platforms, making it possible to use these social networks to identify stress. First, the authors of this article identify a set of stress-related textual, visual, and social attributes from various aspects, and then propose a new hybrid model — a factor graph model in combination with a convolutional neural network for using tweet content and social interaction information to identify stress.

The authors of the article «Generation of a Reliable Online Poll» [1] have developed a tool that will help conduct online surveys using personal information to obtain relevant results.

In the article «Automated Data Collection on Financial Transactions and Social Media» [3], the authors propose to get access to data on social networks in order to get an idea of the general social status of a person, since at the moment the financial reliability of a person is the main factor in approving loans or authorization of credit operations. Today, the person’s “credit rating” is calculated on the basis of the past performance of debt obligations but financial transactions are not the only parameters that can determine the trust in a person. Extracted information from social networks about education, place of work, social status, as well as identifying relatives with credit scores can greatly affect a person’s financial reliability. The authors believe that these factors should be taken into account when calculating a credit score plus additionally read information from SMS messages from the bank about transactions.

Thus, the extracted data from social networks can serve as a basis for a research and development aimed at positive changes in society, including

not only the optimization of various processes but also the improvement of the physical and emotional state of people, which can later reduce the load in the healthcare industry.

Russian legislation in the field of using personal data from social networks

In accordance with the law on the protection of personal data, Federal Service for Supervision in the Sphere of Telecom, Information Technologies and Mass Communications prohibited the free processing of data from open profiles in social networks. Federal Service believes that the extraction of data from social networks violates the law «About Personal Data», because before processing personal information, you must obtain the direct consent of a user because when registering, a user agrees only to the access to the information, which he or she provides on his personal page.

According to the recommendations of the Ministry, personal data is any information relating directly or indirectly to an individual. This list includes last name, first name, middle name, physical address, e-mail, date of birth, telephone number, photo and other data.

Hence, if a company collects, records, stores, depersonalizes, or transmits data, then such actions are considered processing. It turns out that if the company copied to its database the user's information without direct permission from the user, then it violated the information privacy law.

Users may reluctantly allow any organization to have the direct access to their data for uploading, processing, storing and analyzing. This is due to fear of information leaks or tricksters who can use personal information and the results of information for a personal gain. Therefore, access to data collected for processing, storage and analysis must be extremely confidential and must be encrypted in order to avoid illegal use of data. Hence, it is necessary to use reliable data encryption algorithms.

Conclusion

This article describes the types of data that can be extracted from their social networks: Profile data, Social graph and Traffic data.

The options for using a set of data from social networks, described by the authors of scientific articles [1], [2], [3]. Research and development, using information extracted from social networks, can both optimize business processes and provide the basis for positive changes in society.

According to Russian law, those who want to extract data from social networks first need to get consent for an action from users whose data will be extracted. It is also necessary to organize strong encryption and storage of the extracted information so that third parties cannot access it.

To sum up, it is impossible to identify the best way to use the extracted personal data from social networks, because the goals of each research can be completely different and be aimed at improving in different areas of society.

REFERENCES

1. Kodad Mohcine. Generation of a Reliable Online Poll Using the Personals Informations from the Social Networks / Mohcine Kodad, El Miloud Jaara, Mohammed Erramdani // Computer Science and Engineering 2015, 5 (1): 8-13. [Electronic recource]. URL: http://article.sapub.org/10.5923.j.computer.20150501.02.html/). (12.11.2018).

2. Lin Huijie. Detecting Stress Based on Social Interactions in Social Networks / Huijie Lin, Jia Jia, Jiezhong Qiu, Yongfeng Zhang, Guangyao Shen, Lexing Xie, Jie Tang, Ling Feng, and Tat-Seng Chua J. // IEEE Transactions On Knowledge And Data Engineering, September 2017, №9, Vol. 29. [Electronic recource]. URL: https://ieeexplore.ieee.org/abstract/document/7885098/ (07.10.2018).

3. Lohokare Jay. Automated Data Collection for Credit Score Calculation based on Financial Transactions and Social Media / Jay Lohokare, Reshul Dani, Sumedh Sontakke // International Conference on Emerging Trends & Innovation in ICT (ICEI), Pune Institute of Computer Technology, Pune, India, Feb 3-5, 2017. [Electronic recource]. URL: https://ieeexplore.ieee.org/document/7977024/ (02.12.2018).

4. Sasse Robert. A Micro-Economic Perspective on Social Media in Context of the New Economy // Microeconomics and Macroeconomics 2016, 4(2): 56-59. [Electronic recource]. URL: http://article.sapub.org/10.5923.j.m2economics.20160402.03.html/). (01.10.2018).